

DOCUMENT RESUME

ED 200 610

TM 810 118

AUTHOR Ward, William C.
TITLE Evaluation of Videotaped and Live Theatre Auditions.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-80-35
PUB DATE Dec 80
NOTE 19p.
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Evaluation Methods; High Schools; *Talent Identification; Videotape Recordings
IDENTIFIERS *Audition (Theatrical); *Scholars in the Arts Program

ABSTRACT

Theatre auditions by 24 semifinalists in the 1980 Scholars in the Arts program were evaluated under two conditions. Four judges ranked the live auditions, while five evaluated videotapes of the same performance of the high school seniors. The auditions were videotaped in black and white. A single camera was used, fixed at an intermediate distance so as to give a whole-body view. Tapes were mailed to the judges for viewing on playback equipment. Correlations between the two sets of assessments indicated considerable agreement across methods of evaluation, though they were not equivalent to one another. In the present sample, the use of videotapes for final decisions about nominations for awards would have resulted in a completely different group of nominees from those actually chosen. On the other hand, videotapes used as a screening device would have permitted a 60% reduction in those invited to give live auditions with a loss of only one individual who was eventually determined to be among or near the five most accomplished. Despite the limitations of this effort, its results point clearly to the usefulness of videotapes as a screening device. (Author/RL)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED200610

RESEARCH

REPORT

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

RR-80-35

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.



EVALUATION OF VIDEOTAPED AND LIVE THEATRE AUDITIONS

William C. Ward

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Educational Testing
Service

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."



Educational Testing Service
Princeton, New Jersey
December 1980

TM 810118

Copyright © 1980. Educational Testing Service. All rights reserved.

Evaluation of Videotaped and Live Theatre Auditions

William C. Ward

Abstract

Theatre auditions by 24 high school seniors were evaluated under two conditions. Four judges ranked the live auditions, while five evaluated videotapes of the same performance. Correlations between the two sets of assessments indicated considerable agreement across methods of evaluation, though they were not equivalent to one another. Videotapes could be useful as a screening device, serving to limit the number of individuals to be seen in live audition in an adjudication program.

Acknowledgments

Appreciation is due to the staff of the 1980 Scholars in the Arts program for their interest and cooperation, especially to Marc Minardi, John Winterbottom, and Michael Zieky. I want also to thank the five adjudicators, whose names and affiliations are listed in Attachment A. This work was supported by ETS Development funds.

Evaluation of Videotaped and Live Theatre Auditions

In the 1980 Scholars in the Arts program, almost 400 high school seniors submitted applications in theatre. They were evaluated in a two-stage process. Applications and letters of recommendation were used to select 24 semifinalists; the semifinalists were then seen in audition. Five were eventually recommended to the Presidential Scholars Commission to receive awards.

This procedure gives rise to two concerns--one of quality and one of cost. The judges showed excellent agreement with one another in their evaluations of the auditions; thus, there is little reason to question the validity of their choices from among those who reached semifinalist status. However, there is no way to know whether excellent candidates were missed in the first stage of selection, owing to an inability to present their qualifications effectively on paper. And, in fact, within the semifinalist group there was no relation between the final ranking of the auditions and the preliminary ranking of candidates based on applications; the written materials, therefore, do not appear to provide a very good basis for predicting actual performance. In addition, live auditions are expensive. If the program grows as expected in future years, travel costs associated with bringing any substantial percentage of the applicants together for auditions will overwhelm anticipated financing.

Both these concerns might be met in part by the use of videotaped auditions as part of the application procedure. Whatever their flaws, videotape could provide information about performing ability that cannot

be obtained from written materials. If evaluations of tapes should prove sufficiently similar to evaluations based on live auditions, it would be possible to limit severely the number of individuals invited to audition live while still having some assurance that outstanding candidates were not excluded from the final assessment.

Before initiating the use of tapes, however, we should obtain confirmation that they do provide a valid basis for evaluation. This report describes a preliminary examination of the relation between assessments based on live auditions and those based on videotapes of the same performance.

Method

The 24 Scholars in the Arts auditions constitute the performance evaluated. Each student was given four minutes in which to present two selections, unrestricted except that no more than one selection could be an original work. Props were limited and elaborate costuming was discouraged. The auditions were videotaped in black and white. A single camera was used, fixed at an intermediate distance so as to give a whole-body view.

Five adjudicators agreed to view the videotapes; their names and affiliations are given in Attachment A. They were told that the purpose of the evaluation was to determine whether "...we can make effective use of secondary media as a screening device...to select a very small number of applicants--perhaps 2-3% of the total group--for the live auditions from which the award winners will be selected." They were given a roster listing the performers along with the names of their selections, and a second roster on which students were to be ranked "...from 1, most outstanding

in accomplishment and promise, to 24, least outstanding." No further definition of the selection criteria was offered, and no directions were given as to how the tapes should be viewed. Tapes were mailed to the judges for viewing on their own or rented playback equipment.

Results

Three judges complied fully with the ranking instructions. The two remaining preferred to group the candidates, one using four categories, and the other, five. Tied ranks were assigned in these latter two cases, and correlations among all the sets of rankings were obtained. These correlations, shown in the upper portion of Table 1, range from .18 to .59 with a median value of .36. The reliability of an assessment obtained by averaging across five judges is .73 (coefficient alpha). Thus, while there is only moderate agreement among independent assessments based on videotapes, a score derived by averaging over the entire set of judges is sufficiently reliable to merit consideration.

The evaluations of the live auditions are shown in the middle portion of Table 1. Correlations for the four live judges are based on their independent rankings of the candidates prior to discussion and assignment of final audition ranks. The correlations range from .46 to .80, with a median of .63; the reliability of a score obtained by averaging over judges is .88. Thus the judges who viewed the live performance agreed with one another somewhat more strongly than did those who worked from videotapes. There is, however, no basis for concluding that evaluations of live performance are inherently more reliable than those of videotapes--there is no way, with the present data, to distinguish effects due to videotaping from possible differences among the judges, or to take account

of the opportunities which the judges of live performance had had to discuss the candidates while reviewing their application materials.

Relations between the two sets of evaluations are shown in the lower portion of the table. The correlations range from $-.17$ to $.67$, and have a median of $.30$. Thus, on the average, the evaluations of videotaped performance agree with those of live performance to nearly the same degree as they agree among themselves.

Two total scores were obtained, one for videotape evaluations and one for live evaluations, by summing over all relevant judgments. These two sets of scores correlate $.57$ --high enough to indicate considerable agreement across method of evaluation, though not so high as to indicate that those methods are equivalent to one another. The estimated true score correlation between the two sets of rankings is $.70$.

There is some suggestion in the correlations that two of the assessments based on videotapes differ from the remaining three in their relations to the live judgments. Rankings by judges V1 and V4 correlate substantially with the judgments of live auditions (median r 's of $.53$ and $.57$), while rankings by the remaining three videotape judges have lower relations to those judgments (medians of $.30$, $.26$, and $.20$). Factor analysis confirms this difference. Principal axes factor analysis suggests two correlated factors in the judgments ($r = .38$). All the judgments of live auditions along with those by judge V1 define one factor, while those by judges V2, V3 and V5 define a second factor. Evaluations by judge V4 have their principal loading on the second factor along with a secondary loading on the first.

Another way to look at the relations between the two sets of values is to examine differences in rankings received by individual candidates from one kind of evaluation to the other. (The complete set of rankings, including those by individual judges as well as the two total scores, is given in Attachment B.) First, consider the five individuals who were eventually nominated to become award winners. Listed in order of their initial live audition ranks (1, 2, 3, 4, and 6), these individuals were ranked in the videotape assessment 9, 6, 8, 10, and 16. Thus, if videotaped performance had been used to make a final selection of nominees, none of those who were actually nominated would have been among the award winners. On the other hand, if videotape had been employed to select semifinalists for audition, it would have been necessary to audition live only 10 of the 24 to view all but one of those who eventually received nominations (and, in fact, these 10 would have included all but one of the top eight as judged from live auditions). That is, a 60% reduction in the number of individuals auditioned would have resulted in substantially the same group as those who received nominations or were close to doing so, but with the loss of one individual who was eventually placed in the top five.

It is noteworthy that there was one very large discrepancy in the two sets of rankings involving an individual who was near the bottom of the ranks in the live judging (a rank of 21) but who was placed first in the videotape judging. Four of the five videotape assessors agreed that this individual was the most accomplished of those seen (or, for judges who grouped candidates, was one of those in the highest-rated group). Including this case, there were five instances of substantial discrepancies

in the two total rankings (differences of eight or more positions.) Two involved individuals evaluated as more accomplished in the video than in the live judgments, while three involved the reverse.

A number of possible sources of difference between the two sets of evaluations can be hypothesized. One possibility is a fortuitous one: Videotape judges may have been forced to discount differences in diction and projection because of limitations imposed by the relatively poor sound quality of the tapes. The writer's preferred speculation is that those who were given superior rankings in videotape than in live judgments were generally those who displayed a high level of energy and motion and whose movements showed poise and fluidity. Perhaps these two possibilities are related, in that the lack of good information concerning some aspects of performance would force greater reliance on those aspects which are clearly conveyed on the tapes. There is insufficient evidence available to test these hypotheses. Comments by the videotape judges suggested only that those whose videotaped performance received better evaluations than their live performance were perceived as particularly fresh, imaginative, innovative; while those whose live auditions received better evaluations were seen as more conventional in characterization or in choice of selection.

Several of the judges, in addition to comments on individual students, offered reactions to the conditions under which performance was videotaped or to the use of videotape in general. Their comments are summarized in Attachment C.

Discussion

This study has shown that evaluations based on videotaped auditions are related, but not identical, to evaluations of the live auditions from which the tapes are obtained. In the present sample, the use of videotapes for final decisions about nominations for awards would have resulted in a completely different group of nominees from those actually chosen. On the other hand, videotapes used as a screening device would have permitted a 60% reduction in those invited to give live auditions with a loss of only one individual who was eventually determined to be among or near the five most accomplished.

In considering the magnitude of the relation observed between the two sets of values, several factors mitigating against good agreement should be borne in mind. First, the comparison of methods of evaluation was carried out on a select group, involving between six and seven percent of the program applicants. If the use of applications and recommendations to choose semifinalists had any validity at all, this group was more homogeneous in its range of talent than was the pool of applicants from which it was drawn; we would expect stronger relations between the two sets of assessments if they were applied to the larger, more diverse population.

Second, the quality of the videotapes employed may have affected the results. The videotaping had been carried out under restrictive conditions so as not to interfere with the live judging of the auditions; in particular, only one microphone, fixed some distance from the stage, was permitted. The sound track on the tapes was poor, and several judges thought the lighting could have been improved. It is likely that better tapes would be produced if they were seen as an important part of the

evaluation rather than as an adjunct to it. Better tapes could well make the task of evaluation more reliable, and might affect the basis on which evaluations are made---for example, by allowing greater weight to be given to a student's vocal qualities.

Finally, the assessment of videotapes was carried out by judges who had no opportunity to work together to achieve agreement on their bases and standards for evaluation. In contrast, those who judged the live performance had spent many hours working together to select a group of semifinalists. It is likely that a similar opportunity to influence one another's thinking would result in greater agreement among those evaluating videotaped performance, and this in turn would permit the emergence of a stronger relation to judgments of the live auditions.

This study was a preliminary effort. It would be desirable to repeat it with a number of refinements. For example: (1) An entire applicant pool, rather than those preselected on some basis, should be sampled; this would allow an assessment of the relations between live and videotape evaluation across a broader range of talent. (2) The same judges should be involved in both methods of assessment. A design could be developed in which each judge viewed half of a group live and half on videotape; it would then be possible to determine relations between methods while "holding constant" any differences among judges. (3) Judges should be asked to provide systematic evaluations of a number of aspects of a candidate's performance as well as an overall evaluation. We could then seek to discover whether there are particular dimensions of accomplishment which are differentially weighted depending on the medium in which evaluation is conducted.

Despite the limitations of this effort, its results point clearly to the usefulness of videotapes as a screening device. Experience with candidate-produced videotapes and more systematic study along the lines suggested above will be needed, however, before we can be certain as to just how severe a selection is desirable at the screening stage and just how closely videotape and live assessments can be brought into agreement with one another.

Table 1

Correlations Among Rankings by Individual Judges

Correlations Among Rankings of Videotaped Auditions					
	V2	V3	V4	V5	
V1	.38	.21	.42	.18	
V2		.59	.34	.32	
V3			.38	.23	
V4				.51	
Correlations Among Rankings of Live Auditions					
	L2	L3	L4		
L1	.77	.60	.80		
L2		.46	.67		
L3			.60		
Correlations Between Videotaped and Live Audition Rankings					
	V1	V2	V3	V4	V5
L1	.65	.35	.18	.53	.18
L2	.67	.30	.28	.60	.29
L3	.40	.13	.23	.32	-.17
L4	.40	.30	.33	.64	.22

Note: With 24 cases a correlation of .51 is significant at the 1% level of confidence, while one of .41 is significant at the 5% level.

Attachment A

Videotape Adjudicators

Earle R. Gister, Associate Dean
Yale School of Drama/
Yale Repertory Theatre

Professor Arthur Storch, Chairman
Department of Drama
Syracuse University

Professor Miriam Tulin
Department of Drama
Hofstra University

Henry A. Wicke, Jr., Chairman
Department of Theatre
Director, Division of the Arts
The Packer Collegiate Institute

George Wojtasik, Managing Director
Equity Library Theatre

Attachment B

Rankings by Individual Judges and Pooled Over Judges

Live Auditions					Videotaped Auditions					
Pooled	L1	L2	L3	L4	Pooled	V1	V2	V3	V4	V5
1	1	4	5	1	9	7	11	10	9.5	12.5
2	3	2	3	4	6	4	8	19	4	5.5
3	5	1	6	3	8	2	16	6	9.5	12.5
4	4	8	1	5	10	8	5	11	14.5	12.5
5	2	9	4	6	2	10	12	3	4	2
6	8	5	2	7	16	14	18	8	9.5	19.5
7	10	3	15	9	3	9	1	5	4	12.5
8	9	10	10	10	4	5	4	2	14.5	12.5
9	6	13	11	11	12	13	6	21	9.5	12.5
10	11	6	17	12	15	11	19	22	4	12.5
11	13	20	12	2	5	16	7	7	4	5.5
12	15	12	16	8	17.5	18	23	12	4	12.5
13	7	11	21	13	11	1	9	15	14.5	19.5
14	16	18	9	14	24	15	21	23	21	23
15	20	7	14	20	7	6	10	9	14.5	2
16	18	17	8	19	21	23	17	14	21	23
17.5	12	15	23	16	19	19	14	20	21	5.5
17.5	19	19	13	15	17.5	22	3	4	21	19.5
19	17	21	7	23	13	3	13	13	14.5	19.5
20	14	14	22	22	14	12	15	18	14.5	5.5
21	21	16	24	17	1	17	2	1	4	2
22	24	23	18	18	23	20	24	24	21	12.5
23	23	22	20	21	20	24	22	17	21	12.5
24	22	24	19	24	22	21	20	16	21	23

Attachment C

Comments by Judges (Selected and Edited)

On the Use of Videotapes

- V1: The good performers still jumped out, and the poor performers also were obvious. The middle students could have been a tossup for many reasons--fatigue, the video, the lengthy viewing, etc. My sense is that video might prove to be a viable option to live audition if good results are obtained in this evaluation.
- V2: I did not like using the tape. I felt distanced from the event and from the candidates. This did not help me to be objective, it kept me from relying on subjective responses I have learned to trust and use. The TV medium is too cool; it removes the synergistic potential of live theatre. Only in the case of the most dynamic performer was his actual presence felt on the screen. That would be fine if one were judging professionals, but not for novices who have low development of skills.

On These Tapes

- V1: The audio range was limited when an actor moved out of center stage. There was an interfering hum when the operator raised the volume to try to record low voices. The stage floor was a drum, hurting one student who hurled himself about.
- The harsh white light prevented clear facial definition, washing out faces. There was little or no light on the apron of the stage, so that faces were in darkness when an actor was kneeling.
- V2: The quality of the sound was quite poor. For example, one student's physical work was quite free and he used most of the stage, but the stage noise created by the movement made the words inaudible at times.
- The camera moved only across the stage but never at anything closer than a medium shot. Therefore, the features of the candidates were not sharp. I actually could not determine what an actor was experiencing, except through an aural evaluation.
- V3: If the quality of the audio portion of the tape could be improved, it would facilitate the adjudication.
- V4: This series of auditions has many flaws. It may seem that a single long-shot setting approximates the view of the judges as they sit facing the stage. Actually, I found the "technical-tedium" of this frame and the poor sound a very real handicap. It is possible to do a much more acceptable job by learning how and when to use a zoom lens--so that even one camera can give some variety. The TV coverage must be much better--indeed, can be so within a few hours of training.

Attachment C continued

- V5: I believe the process could be improved if after we see a full length shot of the auditioner introducing himself the camera could move in much tighter to a "bust" shot--so that the viewer would be able to see the expressiveness of the auditioner's face.

The quality of the lighting is poor and the sound could be improved.

Other Comments: Method of Rating

- V3: I first viewed the tapes straight through to get an idea of the level of talent and for a first subjective opinion. Then I developed a series of criteria upon which to make a judgment. These criteria, all given equal importance, were: A- Selection of Choice (appropriateness and variety); B- Stage Presence (posture, movement, execution); C- Projection (diction, clarity); D- Control (delivery, energy, pace, attack). Each of the two selections of each candidate was graded individually and then totaled up for a preliminary score. In case of very close scores a third viewing was done and the selection process was made by my own subjective impression.

Other Comments: Choice of Material

- V4: The choice of audition material was, on the whole, quite good. However, adaptation of scenes for solo presentation was sometimes very clumsy and it would be advisable to inform teachers and coaches to give better editing advice in this area. Physical appearance was kept suitably simple--I felt that only one student was actually hindered by costuming.